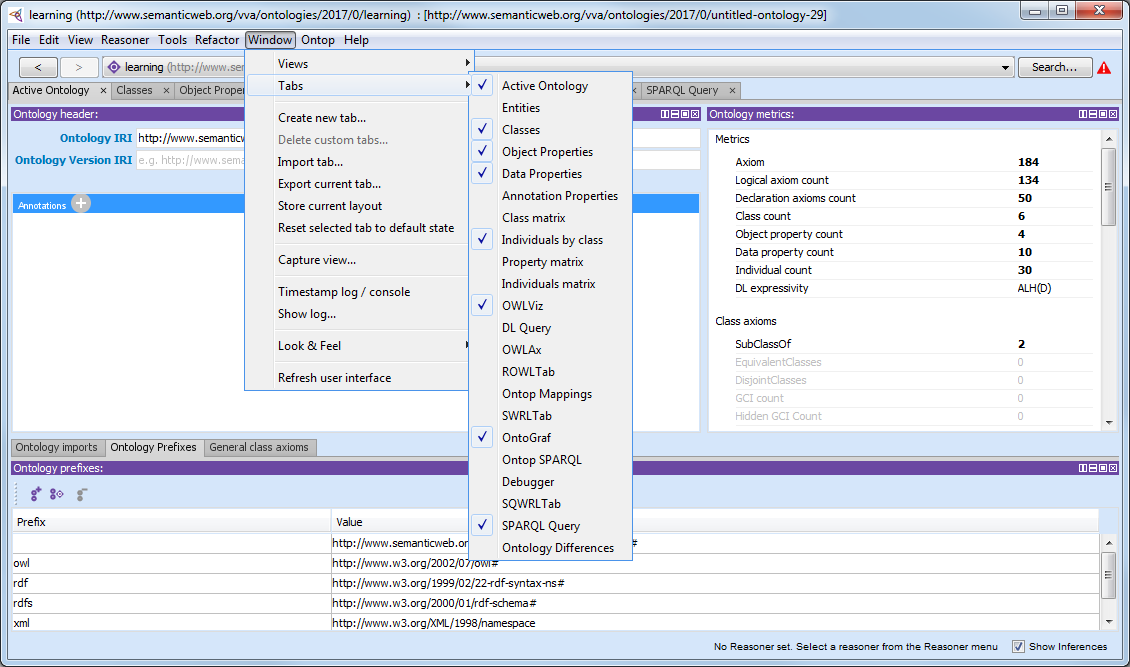
## Лабораторная работа 2

**Онтологии**

# Создание классов

Создание и разработка онтологий с помощью языков RDF/RDFS и OWL, а также выполнение SPARQL-запросов возможны в редакторе Protégé (<http://protege.stanford.edu>).



Назначение вкладок:

- **Active Ontology** – отображение IRI, общих характеристик (количества аксиом, классов, свойств, объектов и т.п.) и аннотации онтологии;

**- Classes** – просмотр и редактирование классов;

- **Object Properties** – просмотр и редактирование свойств-отношений между индивидами;

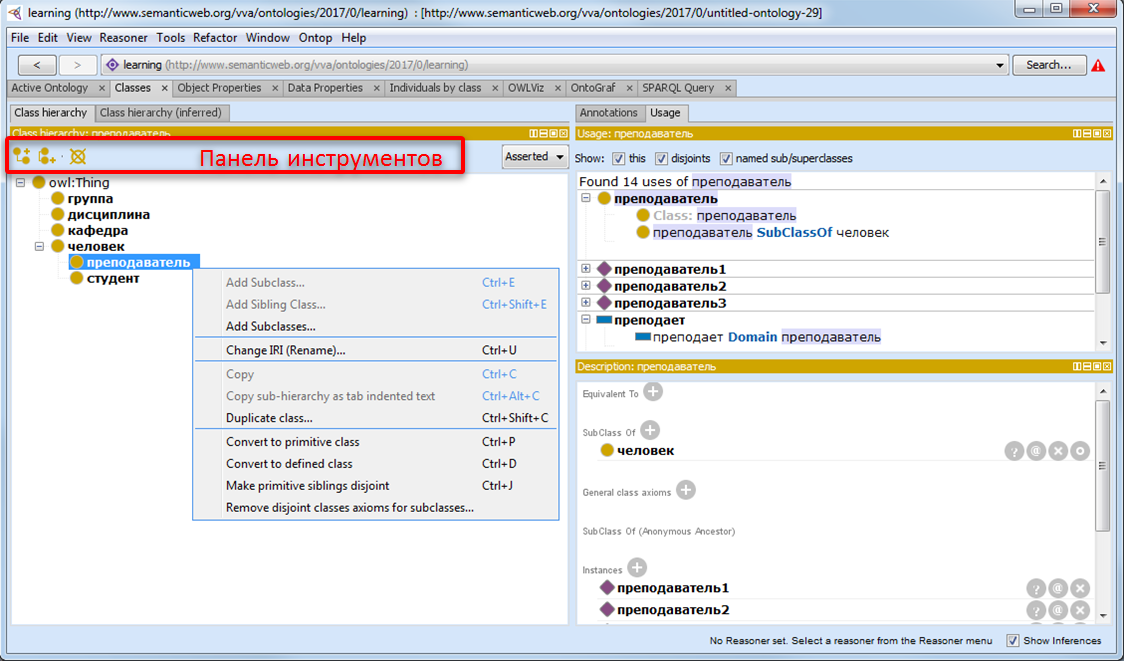
- **Data Properties** – просмотр и редактирование свойств-данных индивидов;

- **Individuals by class** – просмотр и редактирование экземпляров классов (индивидов);

- **OWLVis** и **OntoGraf** – визуализация онтологии в виде графа;

- **SPARQL Query** – выполнение запросов.

Классы создаются на вкладке «Classes». Добавить новый (удалить имеющийся) класс или подкласс можно с помощью кнопок панели инструментов или выбора пункта контекстного меню.



В OWL базовым классом, на основе которого создаются классы онтологии, является класс «owl:Class». Остальные классы по отношению к нему являются дочерними подклассами (англ. Subclass). Классы одного уровня иерархии в Protégé называются родственными (англ. Sibling Class).

На вкладке «Применение» (англ. Usage) для класса, выделенного в иерархии, отображается его связь с родительскими и дочерними классами, его свойства, экземпляры и т.п.

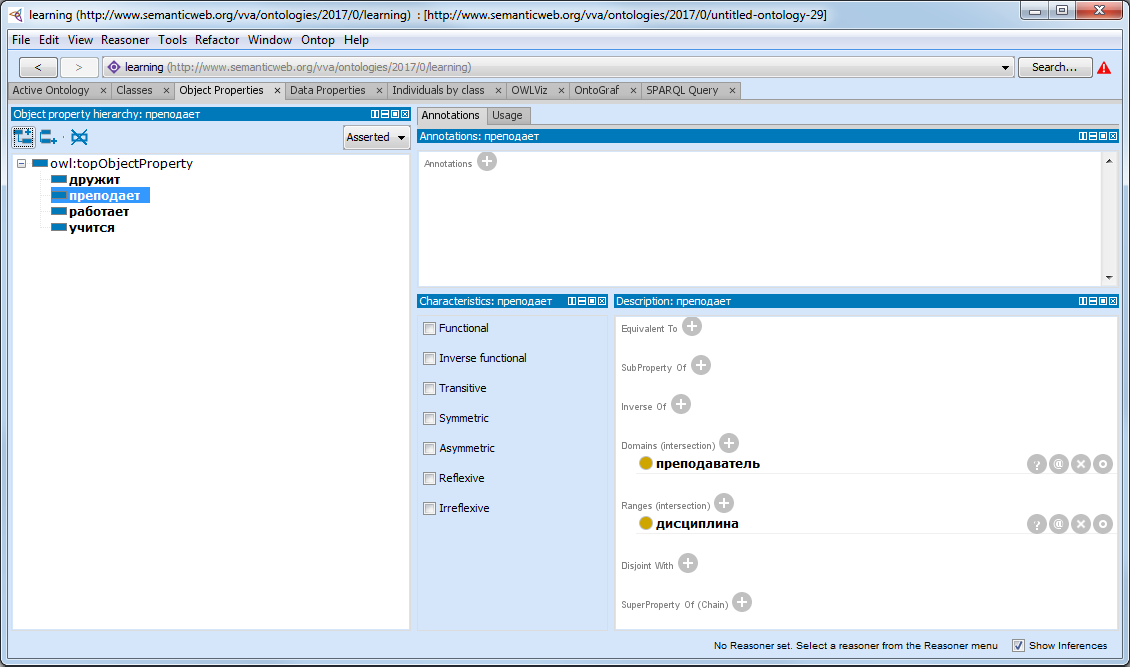
В правой нижней панели «Описание» (англ. Description) можно указать дополнительные характеристики класса. Например, его эквивалентность другим классам (англ. Equivalent To) или невозможность принадлежности экземпляров класса другим классам (англ. Disjoint With) – запрет множественного наследования.

Свойства классов и их экземпляров (предикаты RDF-троек) делятся на два вида:

- **свойства-отношения** задаются на вкладке «Object Properties» и определяют некоторые отношения между двумя индивидами (экземплярами классов), т.е. субъектом и объектом RDF-тройки будут индивиды;

- **свойства-данные** задаются на вкладке «Data Properties» и определяют некоторые фактические характеристики индивидов (экземпляров классов), т.е. субъектом RDF-тройки будет индивид, а объектом значение характеристики в виде строки, числа, даты и т.п.

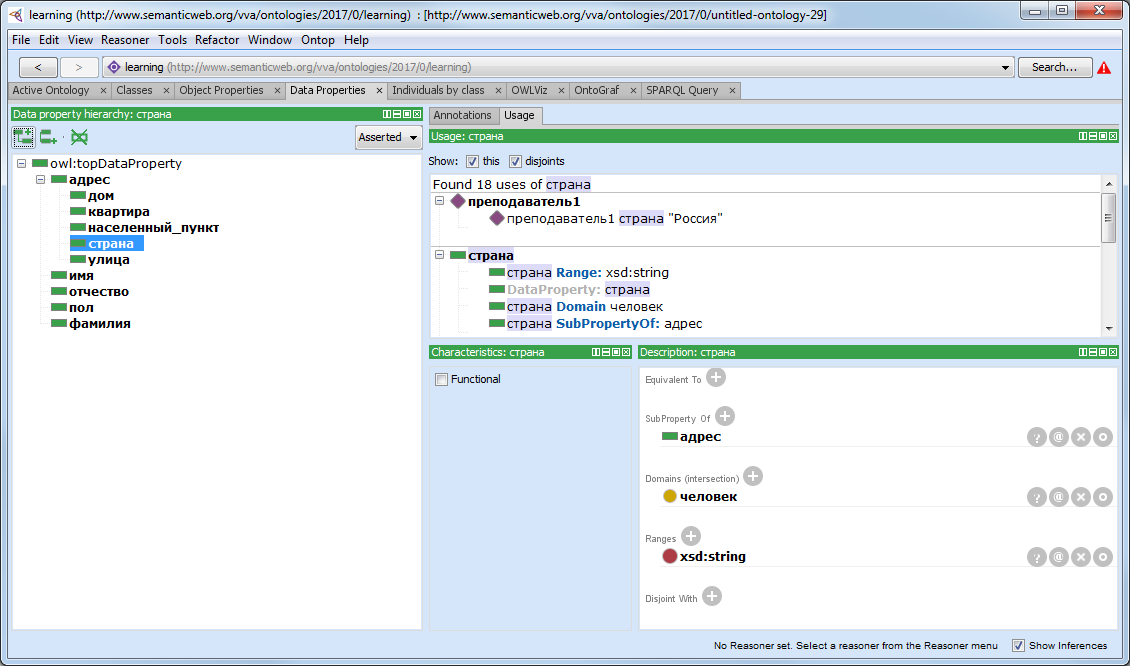
Создание и редактирование свойств-отношений выполняется на вкладке «Object Properties».



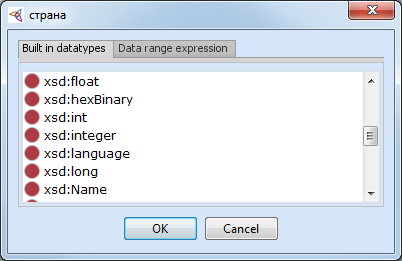
Домен (англ. Domain) указывает, экземпляры каких классов в RDF-тройке при использовании данного свойства будут выступать в качестве субъектов, а диапазон (англ. Range) – в качестве объектов.

На панели «Характеристики» (англ. Characteristics) выбираются дополнительные характеристики свойства: транзитивность (англ.Transitive), симметричность (англ. Symmetric), рефлексивность (англ. Reflexive) и т.п.

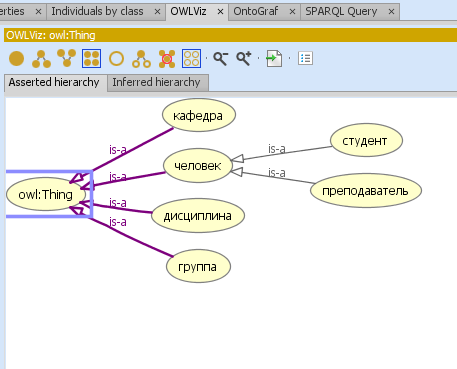
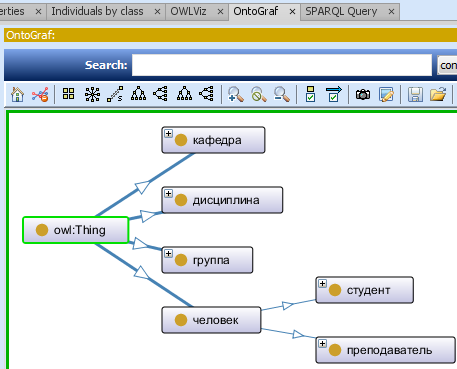
Создание и редактирование свойств-данных выполняется на вкладке «Data Properties».



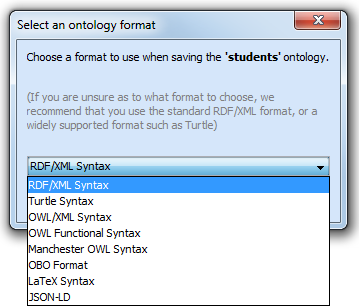
Домен указывает, для экземпляров каких классов данное свойство может быть использовано. Диапазон задает область допустимых значений (тип данных и ограничения), которые можно указывать для данного свойства экземпляру класса. Тип данных выбирается из стандартных типов XML.



Автоматическое построение и отображение графа онтологии и взаимосвязей между классами выполняется при выборе вкладок «OWLViz» и «OntoGraf».

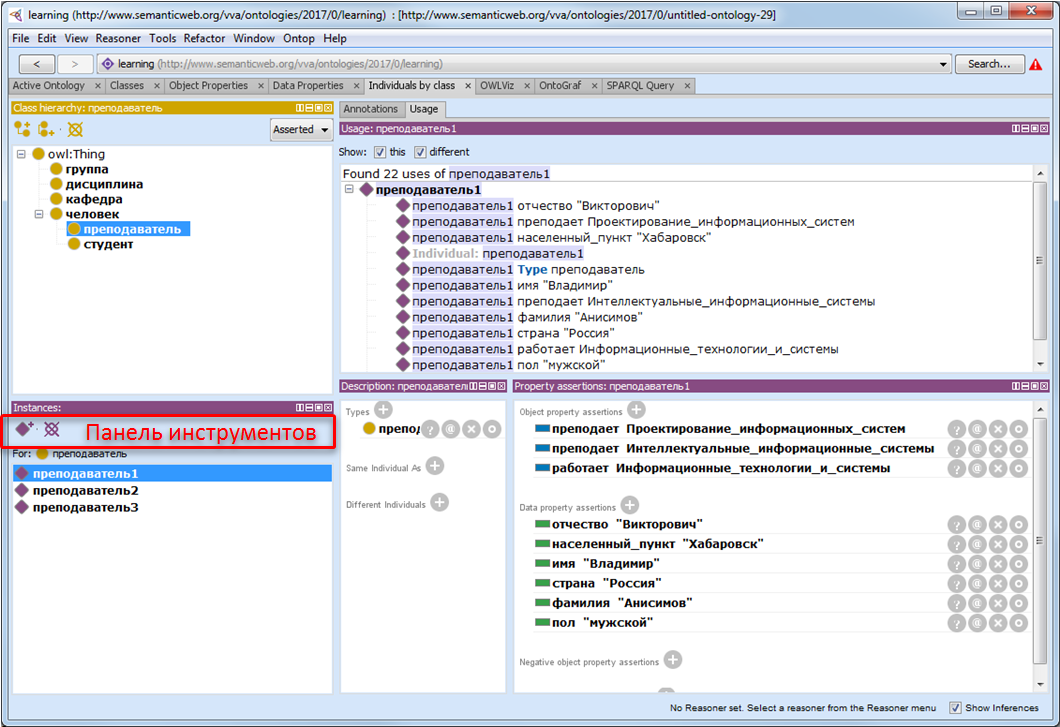
Перед записью на диск (пункт меню «File / Save as …») можно выбрать формат (нотацию) хранения онтологии.



# Создание экземпляров классов

Экземпляры классов в онтологии называются индивидами (англ. Individual). Аналогичным понятием в объектно-ориентированных моделях является объект, но в RDF (OWL) оно зарезервировано за одним из элементов RDF-тройки. В RDF-тройке индивид указывается в качестве субъекта, класс – объекта. Связь между индивидом и классом, представителем которого он является, задается предикатом «rdf:type».

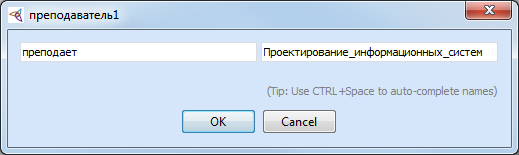
Создание и редактирование индивидов выполняется на вкладке «Individuals by class».



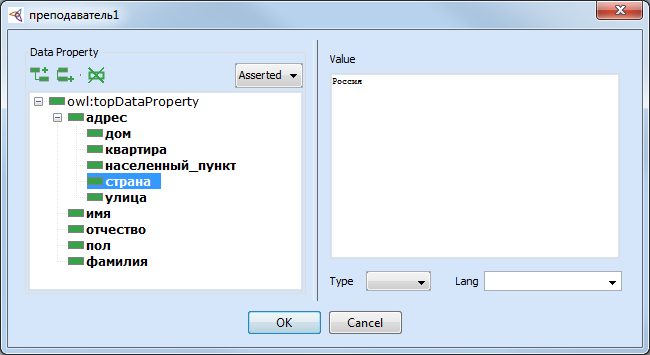
Для добавления индивида необходимо в верхней левой панели «Иерархия классов» (англ. Class hierarchy) выбрать класс, которому он принадлежит, а в панели инструментов нижней левой панели «Экземпляр» (англ. Instances) нажать на первую кнопку. После этого в появившемся окне необходимо указать его имя.

В нижней правой панели «Прикрепленные свойства» (англ. Property assertions) для индивида отображаются его свойства-отношения и свойства-данные.

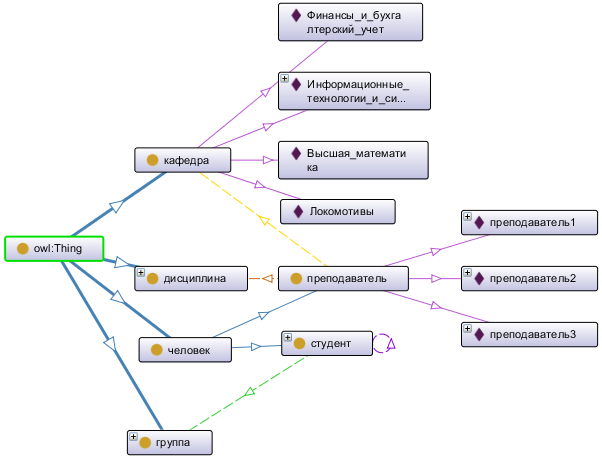
Для их добавления необходимо нажать на иконку серого круга с плюсом после надписи «Object property assertions» («Data property assertions») и в появившемся окне (рис. 2 и 3) определить свойство.



Окно добавления свойства-отношения



Окно добавления свойства-данных



Узлы, помеченные желтыми кружками, являются классами и подклассами онтологии, фиолетовыми ромбами – индивидами. В части дуг графа приняты следующие обозначения:

- синяя сплошная линия – связь между родительским и дочерним классами;

- фиолетовая сплошная линия – связь между классом и индивидом;

- штриховая линия – свойство-отношение между индивидами классов.

# Выполнение SPARQL-запросов в Protégé

По аналогии с базами данных для RDF и OWL онтологических моделей возможно выполнение запросов с целью извлечения знаний, содержащихся в них.

Запрос начинается с ключевого слова «SELECT», после которого указываются:

- переменные, отображаемые в результатах запроса;

- условие «WHERE», задающее ограничения (правила) на выборку результатов.

Переменные, используемые в запросе должны начинаться со знака «?».

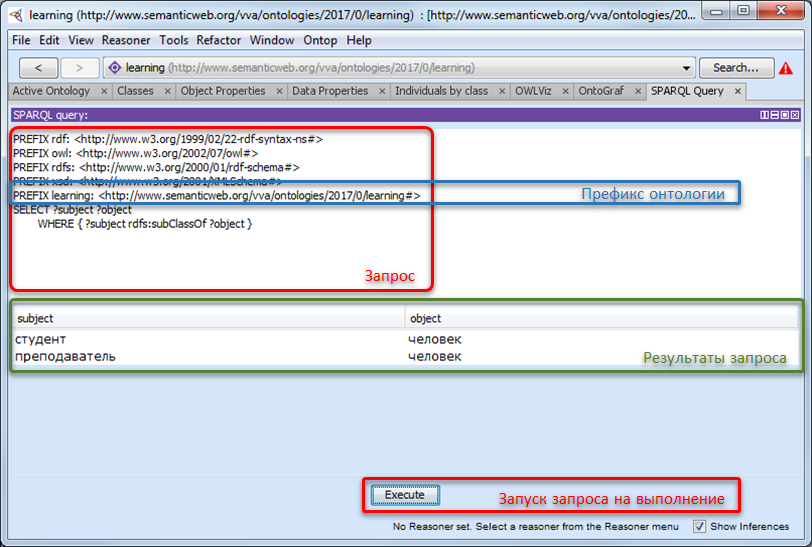
Условие может включать в себя несколько RDF-троек (подусловий), после каждой из которых ставится «.». Считается, что такие подусловия соединены конъюнкцией (логическим И). В условие могут включаться дополнительные конструкции (**графовые шаблоны**):

- FILTER – ограничение на значения переменных;

- OPTIONAL – допущение отсутствия значения для переменной в RDF-тройке;

- UNION – объединение результатов нескольких частей запроса (объединение дизъюнкцией (логическим ИЛИ)).

Выполнение запросов осуществляется на вкладке «SPARQL Query».

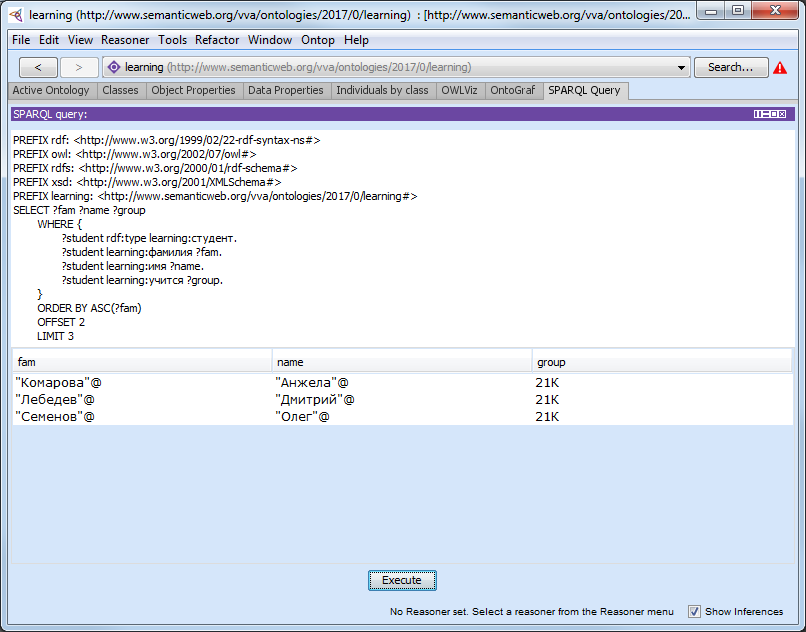


Перед выполнением запросов рекомендуется указать сокращенное имя (префикс) онтологии, чтобы в дальнейшем не было необходимости указания полного IRI для ее элементов (классов, свойств, индивидов).

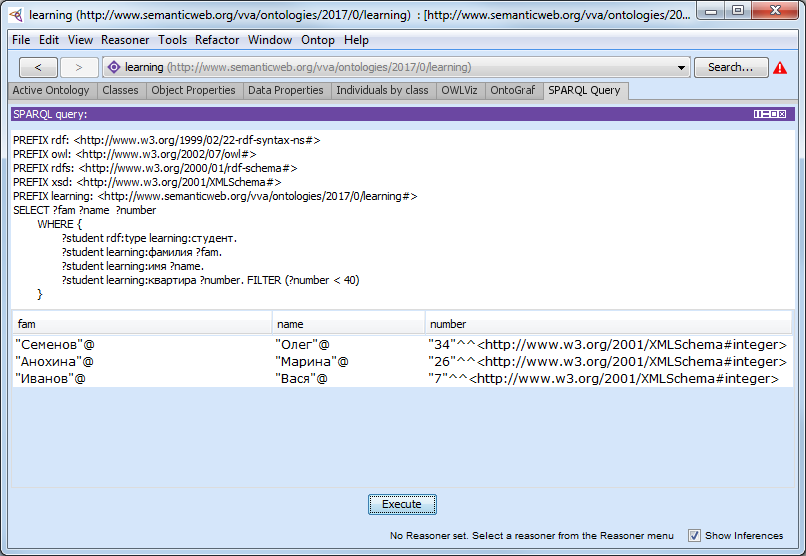
На рис. 1 показан стандартный запрос, отображаемый при первом открытии данной вкладки. Он предписывает вывести все подклассы (переменная ?subject) для всех классов (переменная ?object), определенных в онтологии.

Ниже приводятся несколько запросов с использованием модификаторов и графовых шаблонов.

Показать фамилию, имя и группу 3 студентов (LIMIT 3), пропустив первых 2 (OFFSET 2), в виде отсортированного списка по фамилии (ORDER BY ASC(?fam)).



Показать фамилию, имя и номер квартиры студентов, у которых этот номер меньше 40 (FILTER (?number < 40)).



# Точки доступа Sparql

Точка доступа SPARQL — это служба, поддерживающая протокол запросов SPARQL. Точка доступа позволяет пользователю делать запросы к базе знаний. Сервер обрабатывает запрос и возвращает ответ в некотором, обычно машиночитаемом, формате. Таким образом, точки доступа SPARQL в первую очередь являются API к базам знаний, а представление результатов должно быть реализовано программным обеспечением вызывающей стороны.

Существует много открытых точек доступа SPARQL, например: <https://query.wikidata.org/>

Следующий запрос возвращает список стран и их столиц:

SELECT DISTINCT ?country ?capital ?row

WHERE

{

?country wdt:P31 wd:Q3624078 .

FILTER NOT EXISTS {?country wdt:P31 wd:Q3024240}

OPTIONAL { ?country wdt:P36/rdfs:label ?capital } .

BIND(lang(?capital) as ?row)

filter (?row = 'ru')

}

ORDER BY ?capital

**SELECT, DISTINCT, WHERE и ORDER BY** имеют такой же смысл как и в SQL. Поля в SELECT перечисляются без запятой.

**?country wdt:P31 wd:Q3624078**

это на самом деле запись

Субъект Предикат Объект

То есть выбирается сопоставление “нечто ?country, у которого свойство wdt:P31 равно объекту wd:Q3624078”.

Q3624078 и P31 - уникальные идентификаторы объектов,

wdt и wd - это [префиксы](https://www.mediawiki.org/wiki/Wikibase/Indexing/RDF_Dump_Format#Full_list_of_prefixes) namespace для идентификаторов.

wdt - это описатель какого-то свойства,

P31 – конкретное свойство, заключающееся в том, что ?country является экземпляром чего-либо.

Таким образом, после выполнения строчки

?country wdt:P31 wd:Q3624078

в ?country будет список суверенных государств (свойство wdt:P31 “экземпляр” == значению wd:Q3624078 “суверенное государство”).

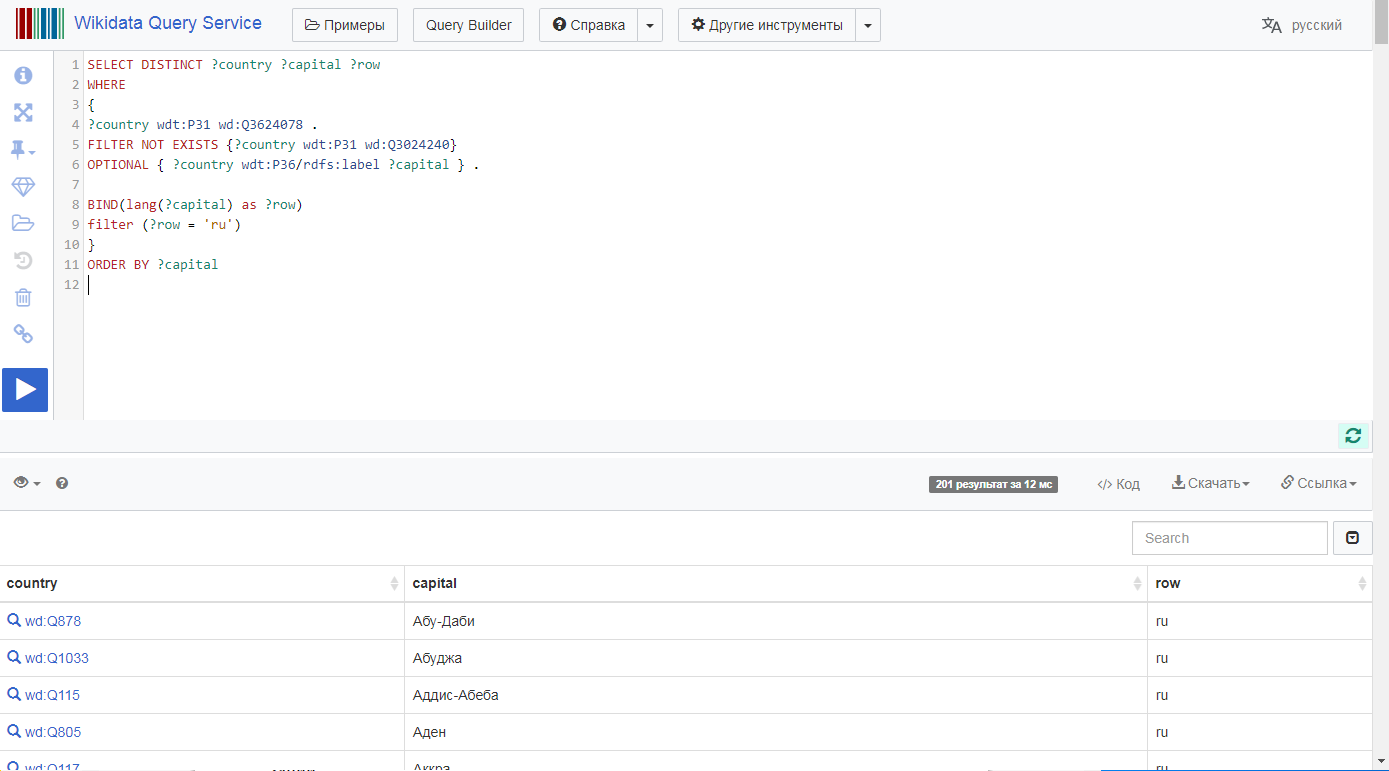
Переходим ко второй строчке в WHERE: исключаем те страны, которые уже не существуют (значение wd:Q3024240). Далее опционально задаём столицу.

**?country wdt:P31 wd:Q3624078 .**

**?country wdt:P36/rdfs:label ?capital .**

Мы хотим найти все существующие страны, у которых есть свойство wdt:P36, которое нужно поместить в переменную ?capital. То есть страны без столиц будут игнорироваться, потому что невозможно заполнить ?capital. Запись через слеш **wdt:P36/rdfs:label**означает вложенное свойство. Нас интересует строка, а не код объекта столицы.

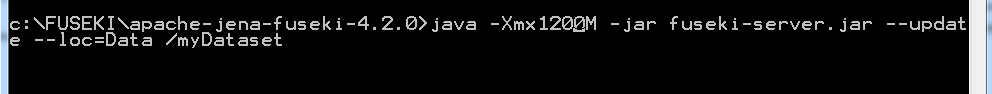
Иногда бывает нужно вывести значение какого-нибудь выражения. В нашем примере это функция lang(?capital), которая возвращает код языка, на котором написано название столицы. С помощью оператора BIND можно записать вычисления в какую-нибудь переменную, а потом использовать это в фильтрах и выводе.



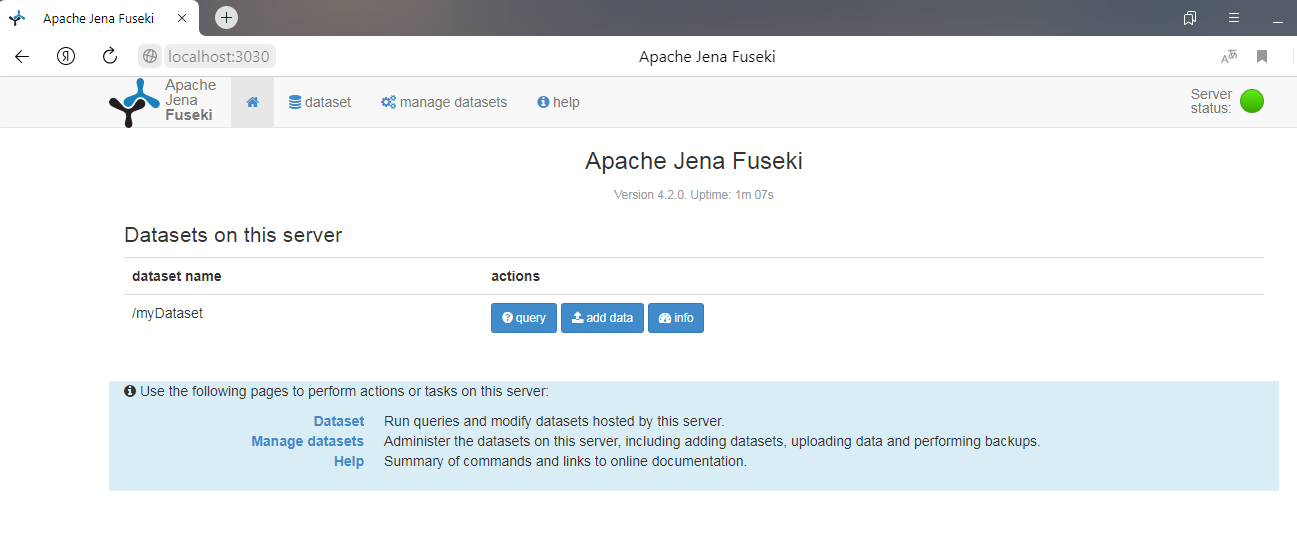
# Создание точки доступа SPARQL

**Jena** - это Java API для RDF.

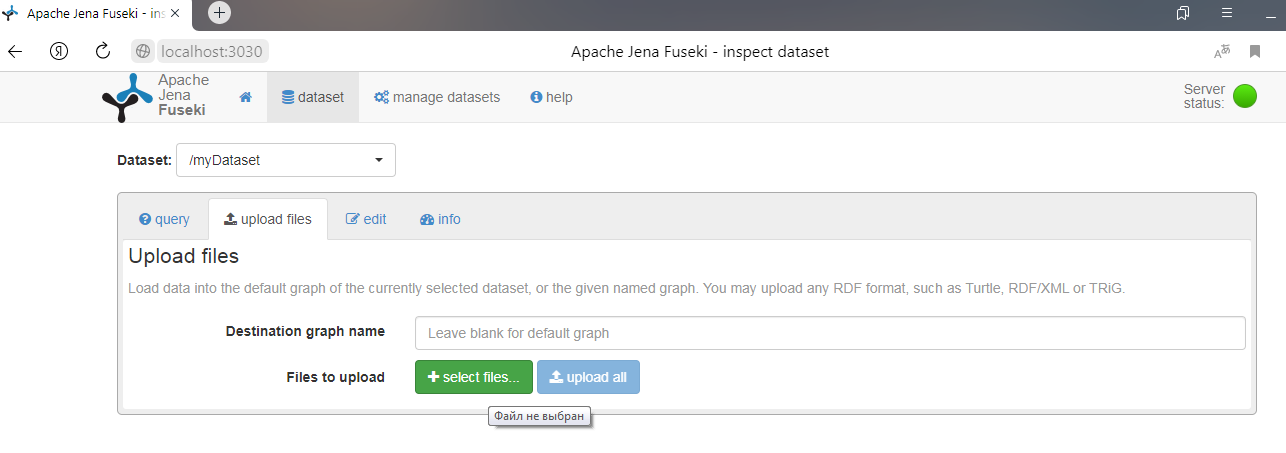
**Fuseki** - это HTTP-интерфейс для данных RDF. Он поддерживает SPARQL для запросов и обновлений. Проект является подпроектом Jena и разработан как сервлет .

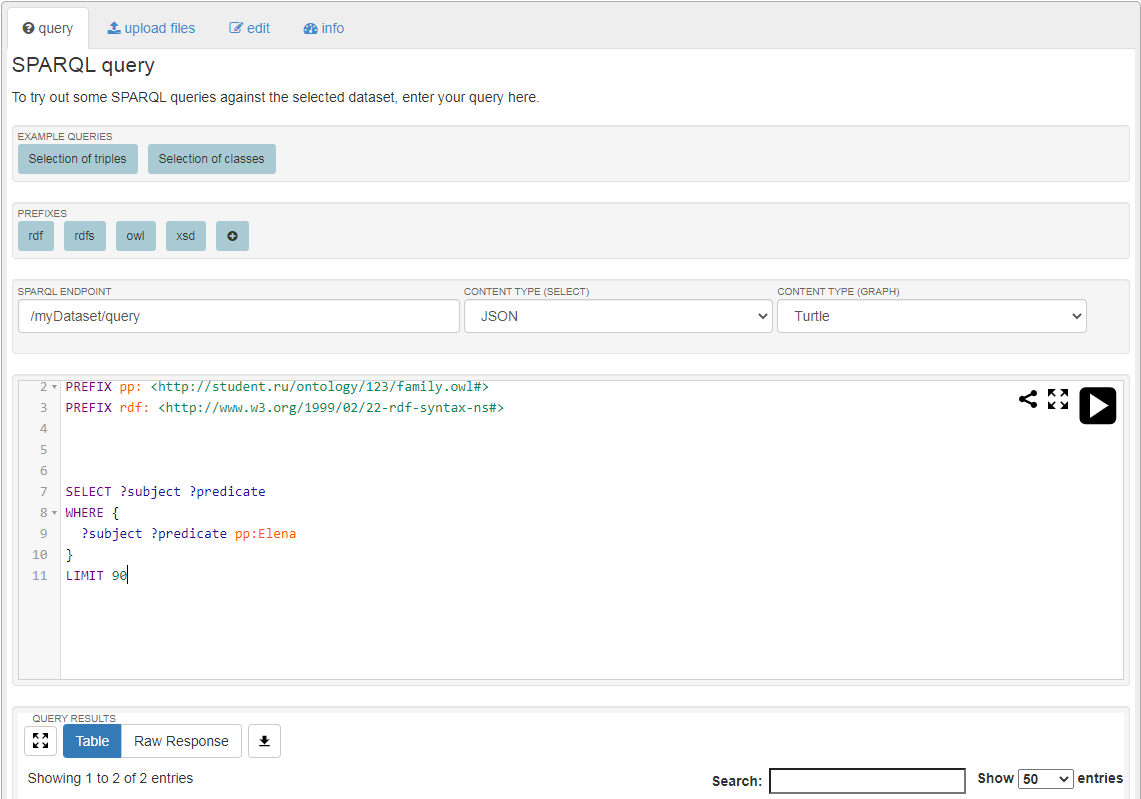


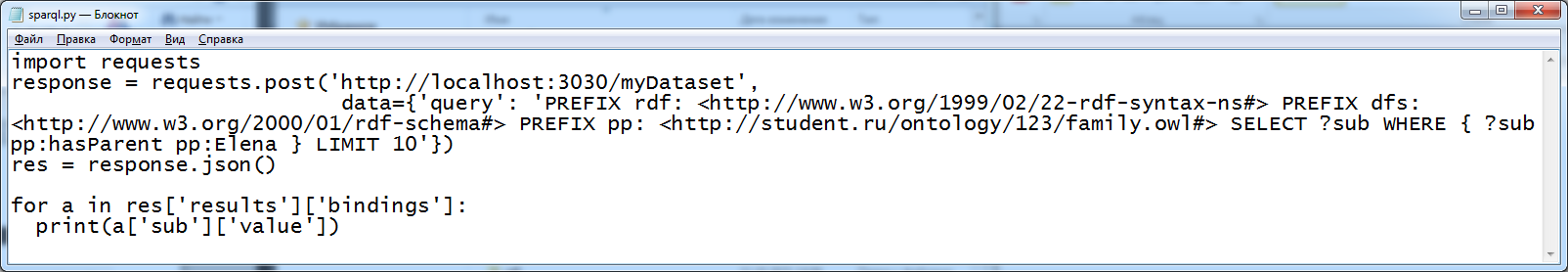
# 



Загрузите RDF файл, созданный в Protege.







# 

# Содержание работы

1. Разработайте онтологию какой-либо предметной области
2. Создайте точку доступа SPARQL. Загрузите созданную базу знаний.
3. Выполните несколько запросов к базе знаний, используя web-интерфейс точки доступа.
4. Создайте python приложение, посылающее запросы точке доступа.